Computing probabilities of very rare events for Langevin processes: a new method based on importance sampling

O. Mazonka[a], C. Jarzyński[b], J. Błocki[a]
(a) Institute for Nuclear Studies, Świerk, Poland
(b) Theoretical Division, Los Alamos National Laboratory, USA
(February 9, 2008)

Abstract

Langevin equations are used to model many processes of physical interest, including low-energy nuclear collisions. In this paper we develop a general method for computing probabilities of very rare events (e.g. small fusion cross-sections) for processes described by Langevin dynamics. As we demonstrate with numerical examples as well as an exactly solvable model, our method can converge to the desired answer at a rate which is orders of magnitude faster than that achieved with direct simulations of the process in question.

PACS: 24.60.Ky, 25.70.Jj, 25.70.-z, 82.20.Fd
Keywords: Langevin processes, low-energy nuclear collisions

INTRODUCTION

Langevin methods offer a powerful tool for the numerical study of low-energy nuclear processes, such as fission and heavy-ion fusion. The evolution of nuclei during such events is typically described using a few collective degrees of freedom, evolving under both conservative and non-conservative forces. The latter, arising from the coupling of the collective variables to the intrinsic nucleonic degrees of freedom, can be modeled by a noisy and a dissipative term in a Langevin description of the collective motion [1]. Once such a stochastic equation of motion has been written down, it is straightforward to numerically simulate the process in question, using a random number generator to supply the noise. By repeating the simulation – with different sequences of random numbers – one obtains independent realizations of the process in question, reflecting the statistical distribution of events occurring during an experiment.

The "direct simulation" method outlined above becomes impractical when studying rare outcomes. For instance, if we are interested in computing the very small cross-section for the fusion of two heavy nuclei, then the vast majority of realizations will end with the nuclei flying apart, and the number of simulations required to obtain even a handful of fusion events may well be prohibitively large. The purpose of the present paper is to propose a general strategy for getting around this problem.

The basic idea which we shall present is essentially a dynamical variant of *importance sampling*, which amounts to gaining information about one probability distribution (a "target" distribution, T), by choosing randomly from another (a "sampling" distribution, S) defined on the same space, and then *biasing* – assigning weights to – the points sampled. The weights are assigned in such a way that the *biased average* of any quantity, over N points drawn independently from S, and the *unbiased average* of that same quantity over N points drawn from T, converge to the same value in the limit of infinitely many samples, $N \to \infty$. If the biased average converges faster with N than the unbiased one, then importance sampling becomes a practical tool for increasing the efficiency of the numerical estimation of the desired average.

In our case, we are interested in Langevin trajectories describing (for instance) the collision of two heavy nuclei, with a very small probability for fusion. Our *target* ensemble, T, is then the statistical distribution of all such trajectories with, say, a given initial center-of-mass energy and impact parameter. The probability of fusion which we wish to compute is defined with respect to this ensemble of trajectories. Our *sampling* ensemble, S, is the distribution of trajectories evolving – from the same initial conditions – under a *modified* Langevin equation, which by design is far more likely to result in fusion. The scheme which we propose involves running a number of simulations with the modified equation of motion (thus obtaining fusion events with good statistics), then computing the desired fusion probability from this data, by biasing each trajectory.

The organization of this paper is as follows. In Section I, we derive the method which we propose in this paper. This section begins with a general discussion of importance sampling, before focusing specifically on its application in the context of Langevin trajectories. In Section II we briefly touch on a number of practical issues associated with the actual implementation of the method. Section III presents numerical results, in which we illustrate the efficiency of the method using a schematic model of nuclear collisions. In Section IV we

analyze how the method behaves for the case of an exactly solvable model, and we conclude in Section V.

We should stress at the outset that, while we shall present our result within the specific context of computing nuclear fusion probabilities, the method we propose can in principle be applied quite generally, whenever one is interested in rare outcomes of processes described by Langevin dynamics. For a different approach to studying rare events (developed in the context of chemical transitions), see the recent work of Chandler and collaborators [2].

I. THEORY

A. Importance Sampling

Importance sampling is based on a very simple idea, embodied by Eq.4 below. Suppose we have some space (ζ-space) on which are defined two normalized probability distributions, $p_S(\zeta)$ and $p_T(\zeta)$, corresponding to "sampling" and "target" ensembles, S and T. Supposing furthermore that $p_S(\zeta) > 0$ whenever $p_T(\zeta) > 0$, let us introduce a *biasing function*

$$w(\zeta) = \frac{p_T(\zeta)}{p_S(\zeta)}, \tag{1}$$

defined at all points ζ for which $p_S(\zeta) > 0$. Now let $\langle \mathcal{O} \rangle_S$ and $\langle \mathcal{O} \rangle_T$ denote the averages of some observable $\mathcal{O}(\zeta)$ over the two distributions:

$$\langle \mathcal{O} \rangle_i \equiv \int d\zeta \, p_i(\zeta) \, \mathcal{O}(\zeta) \qquad , \qquad i = S, T. \tag{2}$$

If we are interested in computing $\langle \mathcal{O} \rangle_T$, then we can do so by repeatedly sampling from the target ensemble T:

$$\langle \mathcal{O} \rangle_T = \lim_{N \to \infty} (1/N) \sum_{n=1}^{N} \mathcal{O}(\zeta_n^T), \tag{3}$$

where $\zeta_1^T, \zeta_2^T, \cdots$ is a sequence of points sampled independently from T. However, it follows from Eqs.1 and 2 above that we can equally well express the desired average as:

$$\langle \mathcal{O} \rangle_T = \langle w\mathcal{O} \rangle_S = \lim_{N \to \infty} (1/N) \sum_{n=1}^{N} w(\zeta_n^S) \, \mathcal{O}(\zeta_n^S), \tag{4}$$

where $\zeta_1^S, \zeta_2^S, \cdots$ is a sequence of points sampled independently from S. Thus, provided we can compute $w(\zeta)$ and $\mathcal{O}(\zeta)$ for any ζ, Eq.4 gives us a prescription for determining the average of $\mathcal{O}$ over the target ensemble T, *using points drawn from the sampling ensemble* S. This prescription becomes a useful tool if a sampling distribution can be chosen for which the rate of convergence with the number of samples (N) is faster when using Eq.4, than when sampling directly from T (Eq.3).

Let us now consider how we might apply importance sampling to the problem which interests us, namely, computing the probability of fusion for two colliding nuclei. For simplicity of presentation, we assume for the moment that the collision process is described by

the evolution of a single collective degree of freedom (e.g. a distance coordinate between the two nuclei), obeying a Langevin equation of the form

$$\frac{dx}{dt} = v_0(x) + \hat{\xi}(t). \tag{5}$$

Here, x is the collective variable, $v_0(x)$ is a "drift" term which embodies the deterministic forces – both conservative and dissipative – acting on the collective degree of freedom, and $\hat{\xi}(t)$ is a stochastic, white noise term:

$$\langle \hat{\xi}(t)\hat{\xi}(t+s)\rangle = D\delta(s), \tag{6}$$

where $\langle\cdots\rangle$ denotes an average over realizations of $\hat{\xi}(t)$, and $D > 0$ is a diffusion constant. [More generally, the number of variables required to specify the instantaneous state of the system will be greater than one: $x \to \vec{x} = (x_1, \cdots, x_d)$. In the case of overdamped motion, this vector specifies the instantaneous configuration of the colliding nuclei. For evolution in which inertial effects are important, the vector $\vec{x}$ will include both configurational variables and their associated momenta. See Section II for elaboration of these and other points.]

Let us assume initial conditions corresponding to a particular value x^0 for the collective variable, and suppose we are interested in the evolution of the colliding nuclei over a time interval $0 \le t \le \tau$. Then a single realization of this process is described by a trajectory $x(t)$, $0 \le t \le \tau$, satisfying $x(0) = x^0$, which obeys Eq.5 for a given realization of the noise term $\hat{\xi}(t)$. Therefore, the statistical ensemble of realizations of $\hat{\xi}(t)$ – together with Eq.5 and the initial condition $x(0) = x^0$ – defines a statistical ensemble of trajectories $x(t)$. By simulating such a process numerically – using a random number generator to provide the noise term – we are, effectively, sampling randomly from this ensemble. Let T ("target") denote the ensemble of trajectories thus defined:

$$T \equiv \left\{ x(t) \,\middle|\, \dot{x} = v_0 + \hat{\xi}\,,\; x(0) = x^0 \right\}, \tag{7}$$

with the time interval $0 \le t \le \tau$ left implicit, here and henceforth.

Given the above Langevin equation (Eq.5), along with the initial condition x^0, we are interested in computing the probability of fusion. Here we will take "fusion" to mean simply that the final point of the trajectory, $x(\tau)$, falls within a given region R along the x-axis:

$$x(\tau) \in R \quad \to \quad \text{fusion} \tag{8}$$

$$x(\tau) \notin R \quad \to \quad \text{no fusion.} \tag{9}$$

Now introduce a functional $\Theta[x(t)]$ which is equal to 1 if the trajectory yields a fusion event, and 0 otherwise.[1] Then the probability of fusion, $P_{\rm fus}$, is just the average of this functional over the ensemble T, defined by Eq.7 above:

$$P_{\rm fus} = \langle \Theta \rangle_T. \tag{10}$$

---

[1] Of course, in our case, the functional $\Theta[x(t)]$ is really just a function of the final point, $x(\tau)$. More generally, however, we could have introduced a criterion for "fusion" which depends on the entire trajectory, in which case $\Theta[x(t)]$ would be a genuine function*al*.

We can compute $P_{\rm fus}$ by sampling randomly from the ensemble T (i.e. repeatedly simulating trajectories evolving under Eq.5), and counting the number of fusion events:

$$P_{\rm fus} \cong P_{\rm fus}^{(N)} = \frac{1}{N}\sum_{n=1}^{N}\Theta[x_n^T(t)], \tag{11}$$

where $x_n^T(t)$ is the n'th of N independent simulations of the process, all launched from x^0. However, since the number of simulations needed to compute $P_{\rm fus}$ to a desired accuracy grows as the inverse of the fusion probability itself (see Eq.29), this "direct sampling" method becomes impractical for very small values of $P_{\rm fus}$. Let us therefore consider a modified version of Eq.5:

$$\frac{dx}{dt} = v_0(x) + \Delta v(x) + \hat{\xi}(t), \tag{12}$$

where the modification consists of adding an extra drift term, Δv, chosen to greatly increase the probability of simulating a fusion event. Let S (for "sampling") denote the statistical ensemble of trajectories $x(t)$ corresponding to *this* equation of motion, with the same initial conditions as above:

$$S \equiv \left\{ x(t) \,\middle|\, \dot{x} = v_0 + \Delta v + \hat{\xi}\,,\; x(0) = x^0 \right\}. \tag{13}$$

Now, for a given trajectory $x(t)$ satisfying $x(0) = x^0$, let $p_T[x(t)]$ denote the probability density that we will obtain this particular trajectory, under the original Langevin equation (Eq.5); and let $p_S[x(t)]$ denote the probability density for obtaining this trajectory under the modified equation (Eq.12). Finally, let w denote the ratio of these two densities[2]:

$$w[x(t)] \equiv \frac{p_T[x(t)]}{p_S[x(t)]}. \tag{14}$$

Then, by Eq.4, we can express the fusion probability $P_{\rm fus}$ (defined with respect to the *original* Langevin equation) as:

$$P_{\rm fus} = \langle w\Theta \rangle_S = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} w[x_n^S(t)] \cdot \Theta[x_n^S(t)], \tag{15}$$

where $x_n^S(t)$ is the n'th of N independent realizations obeying the *modified* Langevin equation. This means that, if we know how to compute w for any trajectory $x(t)$, then we can estimate $P_{\rm fus}$ by running N simulations with the modified Langevin equation, then adding together the weights w for all those trajectories which lead to fusion, and dividing by the total number of simulations, N. In the limit $N \to \infty$, this estimate converges to the correct value of $P_{\rm fus}$.

In the following section we argue that there indeed exists an explicit, easily computable expression for $w[x(t)]$, which renders practical the importance sampling strategy outlined above.

---

[2] Note that while the values of $p_S[x(t)]$ and $p_T[x(t)]$ depend on the *measure* chosen on path space, the ratio $w[x(t)]$ is independent of that measure.

B. Statistical distributions of Langevin trajectories

The original and modified Langevin equations, Eqs.5 and 12, can be represented by the generic equation

$$\frac{dx}{dt} = v(x) + \hat{\xi}(t), \tag{16}$$

where $v = v_0$ in one case, and $v = v_0 + \Delta v$ in the other. As before, given some initial conditions $x(0) = x^0$, let $x(t)$ denote the trajectory evolving from those initial conditions, for a particular realization of the noise term. We are interested in the probability density $p[x(t)]$ for obtaining a particular trajectory $x(t)$. To introduce a measure on the space of all possible trajectories, we discretize the trajectory. Thus, let $(x^0, x^1, \cdots, x^M)$ denote the set of points specifying the state of the system, after time intervals $\delta t = \tau/M$:

$$x^m = x(m\,\delta t) \qquad , \qquad m = 0, \cdots, M. \tag{17}$$

We can then ask for the probability density $p(x^1, \cdots, x^M|x^0)$ that the trajectory passes through the sequence of points $(x^1, x^2, \cdots, x^M)$ at times $\delta t, 2\delta t, \cdots, \tau$, given the initial point x^0. Note that this is a probability distribution in M-dimensional $(x^1, \cdots, x^M)$-space, with x^0 acting as a parameter of the distribution. Let us introduce the notation $\mathbf{X} = (x^0, x^1, \cdots, x^M) = (x^0, \mathbf{Y})$. Then an explicit expression for p is given by [3]:

$$p(\mathbf{Y}|x^0) = (2\pi D\delta t)^{-M/2} \exp -A(\mathbf{X}), \tag{18a}$$

where

$$A(\mathbf{X}) \equiv \frac{1}{2D\delta t} \sum_{m=0}^{M-1} \Big[x^{m+1} - x^m - v(x^m)\delta t\Big]^2. \tag{18b}$$

Eq.18a is strictly speaking valid only in the limit $M \to \infty$ (with τ fixed), although it constitutes a good approximation if $\sqrt{D\delta t}$ is small in comparison with the length scale set by variations in $v(x)$.

Eq.18 is a general expression for the probability density of obtaining a particular discretized trajectory $\mathbf{X}$, launched from x^0. Now let $p_T(\mathbf{Y}|x^0)$ and $p_S(\mathbf{Y}|x^0)$ denote this probability density, for the specific Langevin processes described by Eqs.5 and 12, respectively. Then the ratio between these two probability densities is given by:

$$w(\mathbf{X}) \equiv \frac{p_T(\mathbf{Y}|x^0)}{p_S(\mathbf{Y}|x^0)} = \exp -\Delta A(\mathbf{X}) \tag{19a}$$

$$\Delta A \equiv A_T - A_S, \tag{19b}$$

where A_T and A_S are computed with Eq.18b, using $v = v_0$ and $v = v_0 + \Delta v$, respectively. If we now explicitly write out the expression for ΔA in terms of $(x^0, x^1, \cdots, x^M)$, and consider the limit $M \to \infty$ (with τ fixed), then we arrive at the following result for our biasing function, expressed in terms of an integral over a trajectory $x(t)$ rather than a sum over points along the trajectory:

$$w[x(t)] = \exp -\Delta A \tag{20a}$$

$$\Delta A[x(t)] = \frac{1}{D}\int_0^\tau dt \left(\frac{dx}{dt} - v_0 - \frac{1}{2}\Delta v\right)\Delta v. \tag{20b}$$

Here, dx/dt, v_0 and Δv are evaluated along the trajectory $x(t)$.

C. Computing probabilities of rare events

Combining Eqs.15 and 20, we now have an expression which allows us, in principle, to compute the probability for fusion – defined with respect to the *original* equation of motion (Eq.5) – by running independent simulations with the *modified* equation of motion (Eq.12):

$$P_{\rm fus} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Theta[x_n^S(t)] \cdot \exp -\Delta A[x_n^S(t)]. \tag{21}$$

Here, $x_n^S(t)$ is the trajectory generated during the n'th simulation, using the modified Langevin equation; ΔA is computed for each trajectory[3]; and Θ is, as before, equal to one or zero, depending on whether or not fusion occurred.

For a given set of original and modified Langevin equations, and for a large number N of trajectories simulated under the modified equations, we thus have the following estimator for $P_{\rm fus}$:

$$P_{\rm fus} \cong P_{\rm fus}^{(N)} = \frac{1}{N} \sum_{n=1}^{N} \Theta_n \exp -\Delta A_n, \tag{22}$$

where $\Theta_n \equiv \Theta[x_n^S(t)]$, and $\Delta A_n \equiv \Delta A[x_n^S(t)]$. The estimator $P_{\rm fus}^{(N)}$ converges to the true value $P_{\rm fus}$ in the limit of infinitely many trajectories: $\lim_{N\to\infty} P_{\rm fus}^{(N)} = P_{\rm fus}$.

D. Efficiency analysis

Having derived an estimator for $P_{\rm fus}$ based on the idea of importance sampling, we now consider the question of efficiency. In particular, we establish a specific measure of "how much we gain" by using importance sampling, with a given choice of $\Delta v(x)$.

The validity of Eq.21 does not depend on the form of $\Delta v(x)$. Therefore, for any additional drift term Δv, there will be some threshold value $N_{\Delta v}^*$ such that $P_{\rm fus}^{(N)}$ provides a "good" estimate of $P_{\rm fus}$ for $N \geq N_{\Delta v}^*$. That is, $N_{\Delta v}^*$ is the number of trajectories which we need to simulate (with the modified Langevin equation), in order to determine $P_{\rm fus}$ to some desired accuracy, using the method outlined above. Of course, $N_{\Delta v}^*$ can depend strongly on the form of $\Delta v(x)$. We can thus compare the efficiency of estimating $P_{\rm fus}$, for different drift terms. In particular – since the special case $\Delta v = 0$ is equivalent to computing $P_{\rm fus}$ using the original Langevin equation – let us define the *efficiency gain*, $E_{\Delta v}^G$, associated with a given $\Delta v(x)$, as follows:

$$E_{\Delta v}^G \equiv \frac{N_0^*}{N_{\Delta v}^*}. \tag{23}$$

The numerator is just the number of trajectories needed to accurately estimate $P_{\rm fus}$ by running simulations with the original Langevin equation ($\Delta v = 0$); the denominator is the

[3] While Eq.20b gives a well-defined expression for ΔA, in practice it is more convenient to compute ΔA using the method described in Section II below; see Eqs.32 to 34.

number needed using Eq.22, for a given $\Delta v(x)$. Thus[4] , $E^G_{\Delta v}$ is the factor by which we reduce the computational effort, by making use of importance sampling – again, for a given $\Delta v(x)$.

Let us derive an expression for $E^G_{\Delta v}$ in terms of quantities extracted directly from numerical simulations. For a given additional drift term Δv, let us define

$$f[x(t)] \equiv w[x(t)] \cdot \Theta[x(t)] = \Theta \, \exp -\Delta A. \tag{24}$$

Our method of computing $P_{\rm fus}$ then amounts to computing the average of f, by sampling trajectories $x(t)$ from the ensemble S: $\langle f \rangle_S = P_{\rm fus} \cong P^{(N)}_{\rm fus} = (1/N) \sum_n f_n$, where $f_n = f[x^S_n(t)]$. The statistical error in our result – the expected amount by which $P^{(N)}_{\rm fus}$ differs from $P_{\rm fus}$ – is given by the usual formula for the standard deviation of the mean:

$$\sigma^{(S)}_{P_{\rm fus}} = \frac{\sigma_f}{\sqrt{N}}, \tag{25}$$

where σ_f^2 is the variance of the quantity f over the sampling ensemble,

$$\sigma_f^2 = \langle f^2 \rangle_S - \langle f \rangle_S^2. \tag{26}$$

If we want to compute $P_{\rm fus}$ to a desired relative accuracy r – in the sense that we want the ratio $\sigma_{P_{\rm fus}}/P_{\rm fus}$ (expected error / desired average) to fall below r – then we get the following expression for the minimum number of trajectories needed:

$$N^*_{\Delta v} = \frac{\sigma_f^2}{r^2 P_{\rm fus}^2} = \frac{1}{r^2} \cdot \frac{\langle f^2 \rangle_S - \langle f \rangle_S^2}{\langle f \rangle_S^2}. \tag{27}$$

In other words, if we simulate more than this many trajectories, then we can expect the statistical error in our final estimate of $P_{\rm fus}$ to be no greater than r times $P_{\rm fus}$.

In the case of direct sampling from the target ensemble T, we estimate $P_{\rm fus}$ by computing the average of Θ. The expected statistical error in this average is just

$$\sigma^{(T)}_{P_{\rm fus}} = \left[\frac{\langle \Theta^2 \rangle_T - \langle \Theta \rangle_T^2}{N} \right]^{1/2} = \left[\frac{P_{\rm fus}(1 - P_{\rm fus})}{N} \right]^{1/2}, \tag{28}$$

making use of the fact that $\Theta^2 = \Theta$. By setting this expected statistical error equal to $r P_{\rm fus}$, we get

$$N^*_0 = \frac{1}{r^2} \cdot \frac{1 - P_{\rm fus}}{P_{\rm fus}} \cong \frac{1}{r^2 P_{\rm fus}}, \tag{29}$$

for $P_{\rm fus} \ll 1$. (If we simulate this many trajectories, then the expected number of fusion events is $1/r^2$, and the expected statistical error in this number is $1/r$.)

Combining Eqs.27 and 29, we get the following result for the efficiency gain of our importance sampling method, for a particular choice of $\Delta v(x)$:

---

[4] We are assuming that simulating a single trajectory with the modified Langevin equation takes as much time as simulating one with the original Langevin equation.

$$E^G_{\Delta v} = \frac{N^*_0}{N^*_{\Delta v}} = \frac{P_{\rm fus}(1-P_{\rm fus})}{\sigma_f} \cong \frac{P_{\rm fus}}{\sigma_f} = \frac{\langle f\rangle_S}{\langle f^2\rangle_S - \langle f\rangle^2_S}. \tag{30}$$

Eq.30 gives the efficiency gain of importance sampling, with a particular choice of Δv, in terms of averages which can be estimated from simulations performed under the modified Langevin equation alone (i.e. sampling only from S, not T). An expression for $E^G_{\Delta v}$ in terms of averages estimated using both the original and modified equations of motion, is:

$$E^G_{\Delta v} = \left[\frac{\sigma^{(T)}_{P_{\rm fus}}}{\sigma^{(S)}_{P_{\rm fus}}}\right]^2 = \frac{\langle\Theta\rangle_T - \langle\Theta\rangle^2_T}{\langle f^2\rangle_S - \langle f\rangle^2_S}. \tag{31}$$

We will use these results in Sections III and IV below, to compute the efficiency gain of our importance sampling method for particular examples.

II. PRACTICAL MATTERS

In this section, we discuss a number of practical issues related to the actual implementation of our method.

• At the end of Section I B, we obtained an expression for ΔA as a functional of the trajectory $x(t)$. Since $x(t)$ satisfies Eq.12, we can rewrite Eq.20b as:

$$\Delta A = \frac{1}{2D}\int_0^\tau dt\,(2\hat{\xi} + \Delta v)\Delta v. \tag{32}$$

This expression for ΔA lends itself to a convenient implementation of our method, as follows. When simulating a given trajectory $x(t)$ evolving under Eq.12, we simultaneously integrate the following equation of motion for a new variable $y(t)$, satisfying the initial condition $y(0) = 0$:

$$\frac{dy}{dt} = \frac{\Delta v}{2D}(2\hat{\xi} + \Delta v), \tag{33}$$

for the same realization of the noise term $\hat{\xi}(t)$. (Note that this equation is coupled to the equation of motion for x, since Δv in general depends on x.) Eq.32 then implies that

$$\Delta A = y(\tau). \tag{34}$$

Thus, at the end of the simulation, we use $x(\tau)$ to determine whether or not fusion has occurred, and if so, then we take $\Delta A = y(\tau)$ when assigning the bias $e^{-\Delta A}$ to this event.

• Often (see for instance Section III below), the evolution of our system is such that, once a trajectory $x(t)$ enters the region R which defines fusion, its chance for subsequently escaping that region is negligible: R effectively possesses an absorbing boundary. If this is true for both the original and modified evolution, it becomes convenient to define Δv to be zero everywhere within R. Then, if a trajectory $x(t)$ (evolving under the modified Langevin equation) crosses into R at some time $\tau' < \tau$, we can stop the simulation at that point in time, and take $\Theta = 1$, $\Delta A = y(\tau')$. This saves time, by eliminating the need to continue with the simulation.

• We have, to this point, assumed that the stochastic noise $\hat{\xi}(t)$ is independent of x. That is, D =const. More generally, we might have a diffusion coefficient which depends on the instantaneous configuration of the system: $D = D(x)$, i.e.

$$\langle \hat{\xi}(t)\hat{\xi}(t+s)\rangle_{x(t)=x} = D(x)\delta(s). \tag{35}$$

In this case Eq.18 becomes

$$p(\mathbf{Y}|x^0) = \left\{ \prod_{m=0}^{M-1} \Big[2\pi D(x^m)\delta t\Big]^{-1/2} \right\} \exp -A(\mathbf{X}) \tag{36}$$

$$A(\mathbf{X}) = \frac{1}{2\delta t} \sum_{m=0}^{M-1} \frac{1}{D(x^m)} \Big[x^{m+1} - x^m - v(x^m)\delta t\Big]^2 \qquad . \tag{37}$$

Eq.19 remains unchanged (since the factor multiplying e^{-A} in Eq.36 is the same for p_S and p_T); and Eq.20 changes only in that $1/D$ is brought inside the integral in Eq.20b, with D evaluated along the trajectory $x(t)$. When implementing the method using the additional variable $y(t)$, the only difference is that D is evaluated along $x(t)$ rather than being a constant, in Eq.33.

• Let us now drop the assumption that the system evolves in one dimension. The state of the system is now described by a vector $\vec{x} = (x_1, \cdots, x_d)$, evolving under a set of coupled Langevin equations,

$$\frac{dx_i}{dt} = v_i(\vec{x}) + \hat{\xi}_i(t) \qquad , \qquad i = 1, \cdots, d, \tag{38}$$

where $\vec{v} = \vec{v}_0$ (original equations of motion) or $\vec{v} = \vec{v}_0 + \Delta\vec{v}$ (modified equations of motion). The diffusion coefficient D becomes a symmetric matrix whose elements reflect the correlations between the different components of the stochastic force:

$$\langle \hat{\xi}_i(t)\hat{\xi}_j(t+s)\rangle = D_{ij}\delta(s). \tag{39}$$

[For simplicity, we assume that this matrix is a constant. The generalization to $D_{ij} = D_{ij}(\vec{x})$ is as straightforward as in the one-dimensional case.]

The simplest case of multi-dimensional evolution occurs when the components $\hat{\xi}_i$ are mutually uncorrelated. Then D is a diagonal matrix, and the generalization of Eq.33 is given by:

$$\frac{dy}{dt} = \sum_{i, D_{ii}\neq 0} \frac{\Delta v_i}{2D_{ii}} (2\hat{\xi}_i + \Delta v_i). \tag{40}$$

The sum is taken over values of i for which $D_{ii} \neq 0$, corresponding to those directions along which there is a non-zero stochastic force. Along those directions for which $D_{ii} = 0$, we must have $\Delta v_i = 0$.[5]

If D is not diagonal, then Eq.40 generalizes to the following evolution equation for y:

---

[5] To see this, note that if $D_{ii} = 0$ for a particular value of i, then the equation of motion describing

$$\frac{dy}{dt} = \frac{1}{2}(2\vec{\xi} + \Delta\vec{v})^T D^{-1} \Delta\vec{v}. \tag{41}$$

Eq.41 implicitly assumes that D is invertible, i.e. $\det(D) \neq 0$. If this is not the case, then – first – we must make sure that the projection of $\Delta\vec{v}$ onto the subspace spanned by the null eigenvectors of D is zero. (See the comments following Eq.40.) Assuming this condition is satisfied, we can view Eq.41 as pertaining only to the subspace spanned by the non-zero eigenvectors of D.

The results discussed in this generalization from one-dimensional to multi-dimensional evolution are based on the following generalization of Eq.18:

$$p(\mathbf{Y}|x^0) = \left[(2\pi\delta t)^{d'} \det(D)\right]^{-M/2} \exp -A(\mathbf{X}) \tag{42a}$$

$$A(\mathbf{X}) = \frac{1}{2\delta t} \sum_{m=0}^{M-1} [\vec{x}^{m+1} - \vec{x}^m - \vec{v}(\vec{x}^m)\delta t]^T D^{-1} [\vec{x}^{m+1} - \vec{x}^m - \vec{v}(\vec{x}^m)\delta t]. \tag{42b}$$

As with Eq.41 above, these equations pertain to the $d'(\leq d)$-dimensional subspace spanned by the non-zero eigenvectors of D.

There is an example of multi-dimensional evolution worthy of particular mention. This is the case in which inertial effects are present, i.e. the evolution of the system is not over-damped. The evolution then occurs in the *phase space* of the system, and the equations of motion for $\vec{x} = (\vec{q}, \vec{p})$ are typically of the form:

$$\frac{d\vec{q}}{dt} = I^{-1}\vec{p} \tag{43a}$$

$$\frac{d\vec{p}}{dt} = \vec{F} - \Gamma I^{-1}\vec{p} + \vec{\xi}. \tag{43b}$$

Here, $\vec{q}$ is a vector of variables specifying the configuration of the system; $\vec{p}$ is the vector of associated momenta; I is an inertia tensor; $\vec{F}$ is the vector of conservative forces acting on the system; Γ is a friction tensor; and $\vec{\xi}$ is the vector of stochastic forces, whose associated diffusion tensor D is related to Γ by a fluctuation-dissipation relation. (Typically, I, $\vec{F}$, Γ, and D are all functions of $\vec{q}$.) In this case, the equations of motion for $\vec{q}$ are deterministic (Eq.43a), therefore any additional drift terms must appear only in the *momentum* equations (Eq.43b), as an additional force $\Delta\vec{F}$. The equation of motion for $y(t)$, Eq.41, then pertains only to momentum space (the $\vec{p}$-subspace of phase space).

• Finally, it is often the situation that the dissipative and stochastic forces acting on the collective degrees of freedom depend on the "temperature" of the bi-nuclear system. This is

---

the evolution of x_i is deterministic ($\hat{\xi}_i = 0$). If we now modify that equation by adding an additional non-zero term Δv_i, then any trajectory $\mathbf{x}(t)$ obeying the modified equations of motion will not be a solution of the original equations – for any realization of the noise term $\hat{\xi}(t)$ – and vice-versa (unless Δv_i happens to be zero exactly along the trajectory). This violates the condition stated before Eq.1: in order for the importance sampling to be valid, our modified equations of motion must be capable of generating any trajectory which might be generated by the original equations of motion.

another way of saying that these forces, at time t, depend on the total amount of collective energy which has been dissipated up to that time. Since the energy dissipated, as a function of time, will differ from one realization to the next, it seems we are faced with "memory-dependent" forces, i.e. forces which, at time t, depend not only on the instantaneous state of the system, but also on its *history*, up to time t. An easy way to deal with this situation is simply to expand our list of variables $\vec{x}$ to include a new member, $x_{d+1} = E_{\rm diss}$, denoting the collective energy dissipated. This new variable is initialized at zero, and evolves under an evolution equation which depends on the model used to describe the colliding nuclei. (For instance,

$$\frac{dE_{\rm diss}}{dt} = (I^{-1}\vec{p})^T[\Gamma I^{-1}\vec{p} - \vec{\xi}] \tag{44}$$

in the case of collective evolution described by Eq.43.)

With the addition of this new variable, i.e. $\vec{x} \to (x_1, \cdots, x_d, x_{d+1})$, we now again have a set of (coupled) Langevin equations, in which the drift and stochastic terms depend only on the instantaneous state of the system, $\vec{x}$. We can thus apply the method proposed in this paper, without further modification.

III. NUMERICAL RESULTS

In this Section we describe numerical experiments which we have carried out to test our method, using a simplified model of heavy ion collisions introduced by Świątecki [4]. This model was previously studied by Aguiar *et al* in 1990 [5], using Langevin simulations. For our example, we considered the collision of two ^{100}Zr nuclei. In this mass-symmetric case - for this simple model - the shape of the system is defined by two equal spheres connected by a cylinder. There are two macroscopic ("collective") variables parametrizing the shape: (1) the *relative distance* ρ between the sphere centers, which is the distance s divided by the sum of radii of the two spheres: $\rho = s/2R$; and (2) the *window opening* α, which is the square of the ratio of the cylinder radius to the radius of the sphere: $\alpha = (r_{cyl}/R)^2$.

After some approximations for the potential, dissipation and kinetic energy terms, one obtains the following coupled differential equations for the time evolution of the system (see Ref. [4] for details):

$$\mu\frac{d^2\sigma}{d\tau^2} + \nu^2\frac{d\sigma}{d\tau} + \nu - X = \hat{\xi}_1 \tag{45a}$$

$$\frac{d\nu}{d\tau} - \frac{2\nu + 3\nu^2 - \sigma}{4\nu(\sigma + \nu^2)} = \hat{\xi}_2. \tag{45b}$$

(While Eq.45a has been written here as a second-order stochastic differential equation, in practice we convert it to two first-order equations – one deterministic, one stochastic – by introducing the variable $p_\sigma = \mu\, d\sigma/d\tau$.) Here, the collective coordinates ρ and α are represented by the variables $\nu = \sqrt{\alpha}$ and $\sigma = \rho^2 - 1$; the constant μ is a reduced mass, τ is a reduced time, X is a constant conservative force, and $\hat{\xi}_1$ and $\hat{\xi}_2$ are stochastic forces (Gaussian white noise), related to the dissipative terms by a fluctuation-dissipation relation. (For the ^{100}Zr+^{100}Zr collision, we have $\mu \cong 0.176$ and $X \cong 0.677$.) The evolution of the

colliding nuclei is then represented by a Langevin trajectory in (σ, ν)-space. Fig.1 depicts 30 such trajectories, all starting from a configuration of two touching spheres ($\sigma = 0, \nu = 0$), with a center-of-mass energy equal to 0.8 MeV above the interaction barrier. This energy is about 2.5 MeV below the "extra push" energy, so most of the trajectories (28 of them) lead to reseparation of the system (fission), and only two trajectories lead to a compound nucleus (fusion).

From Fig.1 we have the following picture of the physical process occuring, in the context of this simplified model: first the window opening between the two nuclei grows rapidly; then around a saddle point, at $(\sigma, \nu) \sim (0.0, 0.6)$, the combination of deterministic and stochastic forces determines the ultimate fate of the nuclei, either fusion or reseparation; and finally the system evolves toward its destiny, with σ decreasing in the case of fusion, or increasing with reseparation. This suggests that, if we are to add an extra drift term to increase the likelihood of fusion, then it would be best to localize such a term in the vicinity of the saddle point. We will think of such a term as a *force*, pushing the system toward fusion. We have chosen an additional force along the (negative) σ direction, whose strength is a Gaussian function of (σ, ν), with a peak at $(0.0, 0.6)$. This leads to the following *modified* Langevin equations of motion:

$$\mu \frac{d^2\sigma}{d\tau^2} + \nu^2 \frac{d\sigma}{d\tau} + \nu - X = \hat{\xi}_1 - \Lambda \exp\left\{ -\frac{\sigma^2 + (\nu - 0.6)^2}{0.02} \right\} \tag{46a}$$

$$\frac{d\nu}{d\tau} - \frac{2\nu + 3\nu^2 - \sigma}{4\nu(\sigma + \nu^2)} = \hat{\xi}_2 \quad , \tag{46b}$$

with Λ an adjustable parameter. Fig.2 schematically shows the region around the saddle point, where the additional force pushing the system toward fusion is localized. Two deterministic trajectories (evolving under the original equations of motion, but without the stochastic terms) are also shown, to guide the eye. One of these was launched with an energy of 1 MeV above the barrier (leading to reseparation), the other one at 5 MeV above the barrier (leading to fusion).

To compare our importance sampling method to direct simulation of the original process, we first chose to compute the probability of fusion for trajectories starting with an energy 0.2 MeV above the barrier. This probability is on the order of 10^{-3}, considerably less than that for the case shown in Fig.1 (0.8 MeV above the barrier). We ran 10^4 independent trajectories under both the original and the modified Langevin equations, Eqs.45 and 46, respectively, and kept a running tally of the probablity of fusion, as computed by the two methods (Eq.11 for the case of direct simulation, Eq.22 for importance sampling). We took the strength of the additional force to be $\Lambda = 0.3\mu$ in these simulations. Fig.3 illustrates the difference between the rates of convergence of the two methods. The plots show the estimates $P_{\rm fus}^{(N)}$ as computed by each method, as a function of number of trajectories simulated, N. It is clear that the importance sampling (broken line) converges much faster than direct simulation (solid line): after about 5000 trajectories, the former has converged very close to its asymptotic value, whereas the latter is still "jumping around" after 10000 trajectories.

The sawtooth pattern exhibited by the direct simulation estimate is typical of the situation in which rare events contribute disproportionally to an ensemble average: in this particular case, only 15 of the 10000 "direct" trajectories (simulated under Eq.45) lead to fusion; each of these events causes a sudden jump in the $P_{\rm fus}^{(N)}$, followed by a gradual de-

cline as non-fusion events accumulate. By contrast, under the modified Langevin equations, Eq.46, about 25% of the 10000 trajectories lead to fusion, resulting in smoother and faster convergence.

In Figs.4 and 5 we show excitation functions – fusion probability plotted against center-of-mass energy above the barrier – as computed by both direct simulation and importance sampling, with the same additional force as used in Fig.3. Each point was obtained using 1000 trajectories, and the result is displayed with error bars, as estimated from the numerical data. The solid line represents an analytical formula which closely approximates the fusion probability over the region shown.[6] Again we see that, for approximately the same computational effort, our importance sampling method gives significantly better results than direct simulation. For the point corresponding to 0.5 MeV above the barrier, the error bar in Fig.4 is about 5.5 times bigger than that in Fig.5. The efficiency gain of the importance sampling approach is therefore about 30 ($\sim 5.5^2$, see Eq.31) in this case: we would need to launch about 30×10^3 trajectories evolving under the original Langevin equation to get the same degree of accuracy obtained in Fig.5 with 10^3 trajectories.

The gain in efficiency becomes more dramatic when we go to very small probabilities. To show this we considered the reaction ^{110}Pd+^{110}Pd ($\mu \cong 0.174$, $X \cong 0.794$), for which the extra push energy is 25.5 MeV. Launching 250000 trajectories with an initial center-of-mass energy of 1 MeV above the barrier, we obtained a probability of fusion $P_{\rm fus} = (6.970 \pm 0.268) \times 10^{-13}$. This was computed using our importance sampling method, with an additional force corresponding to $\Lambda = 1.9\mu$; about 88% of the trajectories evolving under the modified Langevin equation went to fusion. Using Eq.30, our result gives an efficiency gain of $E^G_{\Delta v} = 3.5 \times 10^9$! We cannot compare our estimate of $P_{\rm fus}$ directly to an estimate obtained from simulating with the original Langevin equation, since we would need to run $\sim 10^{12}$ trajectories to have a decent chance of observing even a single fusion event. Importance sampling is indispensible in this case: we could not have calculated $P_{\rm fus}$ using direct simulations.

IV. AN EXACTLY SOLVABLE MODEL

In addition to the numerical results of the previous section, it is instructive to consider an exactly solvable model. Consider a particle in two dimensions which falls at a uniform rate v_z from a height h, while experiencing random "kicks" in the horizontal direction:

$$\dot{z} = -v_z \qquad , \qquad \dot{x} = \hat{\xi}(t) \ , \tag{47}$$

where $\hat{\xi}(t)$ represents white noise corresponding to a diffusion constant of unit magnitude: $\langle \hat{\xi}(t)\hat{\xi}(t+s)\rangle = \delta(s)$. We assume that the initial horizontal location is zero, i.e. $x(0) = 0$, and are interested in the horizontal location of the particle when it hits the "ground" ($z = 0$),

[6] This was obtained by running a very large number of simulations for different values of energy above the barrier, and then fitting the results to an exponential multiplied by a second-order polynomial. The expected error associated with the curve itself is everywhere smaller than the smallest of the error bars shown in Fig.5.

at time $\tau = h/v_z$. The motion in the x-direction is a Wiener process, whose solution is a Gaussian distribution with a variance growing linearly with time. Let us suppose we are interested in the probability that the particle will end to the right of the fixed point x_0, i.e. $x(\tau) > x_0$. Since the ensemble distribution of $x(\tau)$-values is a Gaussian (with variance equal to τ), this probability is given in terms of an error function:

$$P[x(\tau) > x_0] \equiv F(x_0) = \frac{1}{2}\Big[1 - \mathrm{erf}(x_0/\sqrt{2\tau})\Big]. \tag{48}$$

For large values of x_0, this probability dies off very rapidly, $P \sim e^{-x_0^2/2\tau}/x_0$, therefore very many simulations would be needed to compute P to some desired relative accuracy r.

Let us now consider an additional force in the form of a constant horizontal "wind", pushing the particle in the direction of the point x_0. The modified equations of motion are then

$$\dot{z} = -v_z \qquad , \qquad \dot{x} = \mathrm{w} + \hat{\xi}(t) \ , \tag{49}$$

where w is the strength of the wind. From Eq.32, it follows that $\Delta A = \mathrm{w}x(\tau) - \mathrm{w}^2\tau/2$. From Eq.27, we can then compute the minimum number of simulations necessary to obtain P to a relative accuracy r, using importance sampling, for a particular value of wind strength:

$$N^*_{\mathrm{w}} = \frac{1}{r^2}\left\{e^{\mathrm{w}^2\tau}\frac{F(x_0 + \mathrm{w}\tau)}{F^2(x_0)} - 1\right\}. \tag{50}$$

The efficiency gain is then:

$$E^G_{\mathrm{w}} \equiv \frac{N^*_0}{N^*_{\mathrm{w}}} = \frac{F(x_0) - F^2(x_0)}{e^{\mathrm{w}^2\tau}F(x_0 + \mathrm{w}\tau) - F^2(x_0)}. \tag{51}$$

For the case $x_0 = 3.0$, $\tau = 1.0$ ($P = 1.35 \times 10^{-3}$), we have plotted efficiency gain as a function of wind strength, in Fig.6. We see that the optimal wind strength (at which we obtain maximal efficiency) is $\mathrm{w}_{opt} = 3.157$. For this value of w, the number of trajectories needed to estimate P using direct simulation of the original Wiener process, is about 220 times the number needed to estimate P (to the same degree of accuracy) with importance sampling.

It is interesting to note that even for $x_0 = 0$ (for which half the trajectories fall to the right of x_0), we gain efficiency by using importance sampling. In this case, for $\tau = 1$, the maximal efficiency gain (about 1.8) is achieved with a wind strength $\mathrm{w} = 0.6125$, as shown in Fig.7.

The maximal efficiency gain grows as the probability P becomes small. Using an asymptotic expression for the error function, we have

$$F(x_0) \to \left(\frac{\tau}{2\pi}\right)^{1/2}\frac{1}{x_0}\exp(-x_0^2/2\tau) \qquad (\tau \text{ fixed}\,,\, x_0 \to \infty), \tag{52}$$

from which it follows that

$$N^*_w \to \frac{x_0}{r^2}\exp\Big[(x_0 - \mathrm{w}\tau)^2/2\tau\Big]. \tag{53}$$

This formula nicely encapsulates the dramatic efficiency gain achieved for small values of P (i.e. large x_0). Without importance sampling, i.e. setting w= 0, the number of trajectories needed grows exponentially in x_0^2: $N_0^* \sim \exp(x_0^2/2\tau)$ (dominant contribution). However, using importance sampling, with the optimal wind value ($\mathrm{w}_{opt} = x_0/\tau$), the number needed grows linearly with x_0: $N_{\mathrm{w}}^* \sim x_0$. Thus,

$$N_{\mathrm{w}}^* \sim \sqrt{\ln N_0^*}. \tag{54}$$

Even when N_0^* is tremendously large, N_{w}^* may remain modest (for $\mathrm{w} = \mathrm{w}_{opt}$). For instance, for $x_0 = 6$ and $\tau = 1$ we have $P[x(\tau) > x_0] = .99 \times 10^{-9}$. One would need $\sim 10^{12}$ direct simulations to compute this probability to 10% accuracy. Using importance sampling, one can achieve the same accuracy with fewer than 1000 trajectories. Numerical results (with $\mathrm{w} = 6.0$) gave us $P[x(\tau) > 6.0] = (1.04 \pm 0.10) \times 10^{-9}$ after only 700 trajectories.

V. CONCLUSIONS

In this paper we have developed a method for computing the probabilities of rare events – exemplified by the fusion of two nuclei – for processes described by Langevin dynamics. The method, based on the idea of importance sampling, is straightforward to implement, quite general, and can lead to a very large increase in computational efficiency. For these reasons we believe it represents a very practical tool for using numerical simulations to compute small probabilities. Indeed, with our method, we were easily able to estimate a fusion probability, within a schematic model of nuclear collisions (see the end of Section III), that would have been essentially impossible to estimate from direct simulations of the process in question. We see every reason to expect similar results when combining the method with more realistic semiclassical models of nuclear dynamics.

ACKNOWLEDGMENTS

The authors wish to acknowledge stimulating conversations with Janusz Skalski and Władek Świątecki. This research grew out of conversations among the authors at the XXV Mazurian Lakes School of Physics (Piaski, Poland, 1997), and was continued during reciprocal visits to the Institute for Nuclear Studies in Poland, and Los Alamos National Laboratory in the United States. Financial support for this work came from the Polish-American Maria Skłodowska-Curie Joint Fund II, under project PAA/NSF-96-253. J.B. also acknowledges support from the Committee of Scientific Research of Poland (KBN Grant No. 2-P03B-143-10).

REFERENCES

[1] See, for instance, P.Fröbrich and I.I.Gontchar, Phys.Rep.**292**, 131 (1998), and references therein.

[2] C.Dellago *et al*, J.Chem.Phys.**108**, 1964 (1998); F.S.Csajka and D.Chandler, "Transition pathways in a many-body system: application to hydrogen-bond breaking in water", J. Chem. Phys., in press; P. Bolhuis, Ch. Dellago and D. Chandler, "Sampling ensembles of deterministic transition pathways", Faraday Discussion, in press.

[3] F.W.Wiegel, *Introduction to Path-Integral Methods in Physics and Polymer Science* (World Scientific, Philadelphia, 1986). Original references are: S.Chandrasekhar, Rev.Mod.Phys.**15**, 1 (1943); L.Onsager and S.Machlup, Phys.Rev.**91**, 1505 (1953) and Phys.Rev.**91**, 1512 (1953).

[4] W.J.Świątecki, Physica Scripta **24**, 113 (1981).

[5] C.E.Aguiar, V.C.Barbosa, and R.Donangelo, Nucl.Phys.A **514**, 205 (1990).

FIGURES

FIG. 1. Thirty trajectories simulated using the schematic model of nuclear collisions, Eq.45. The system is ^{100}Zr+^{100}Zr, at 0.8 MeV above the interaction barrier. Two trajectories lead to fusion; the rest to reseparation.

FIG. 2. The circle indicates the saddle point region of the potential energy, and the arrows show the direction of the extra force chosen to push the system toward fusion. Also shown are two (deterministic) trajectories, one ending in fusion, the other in reseparation.

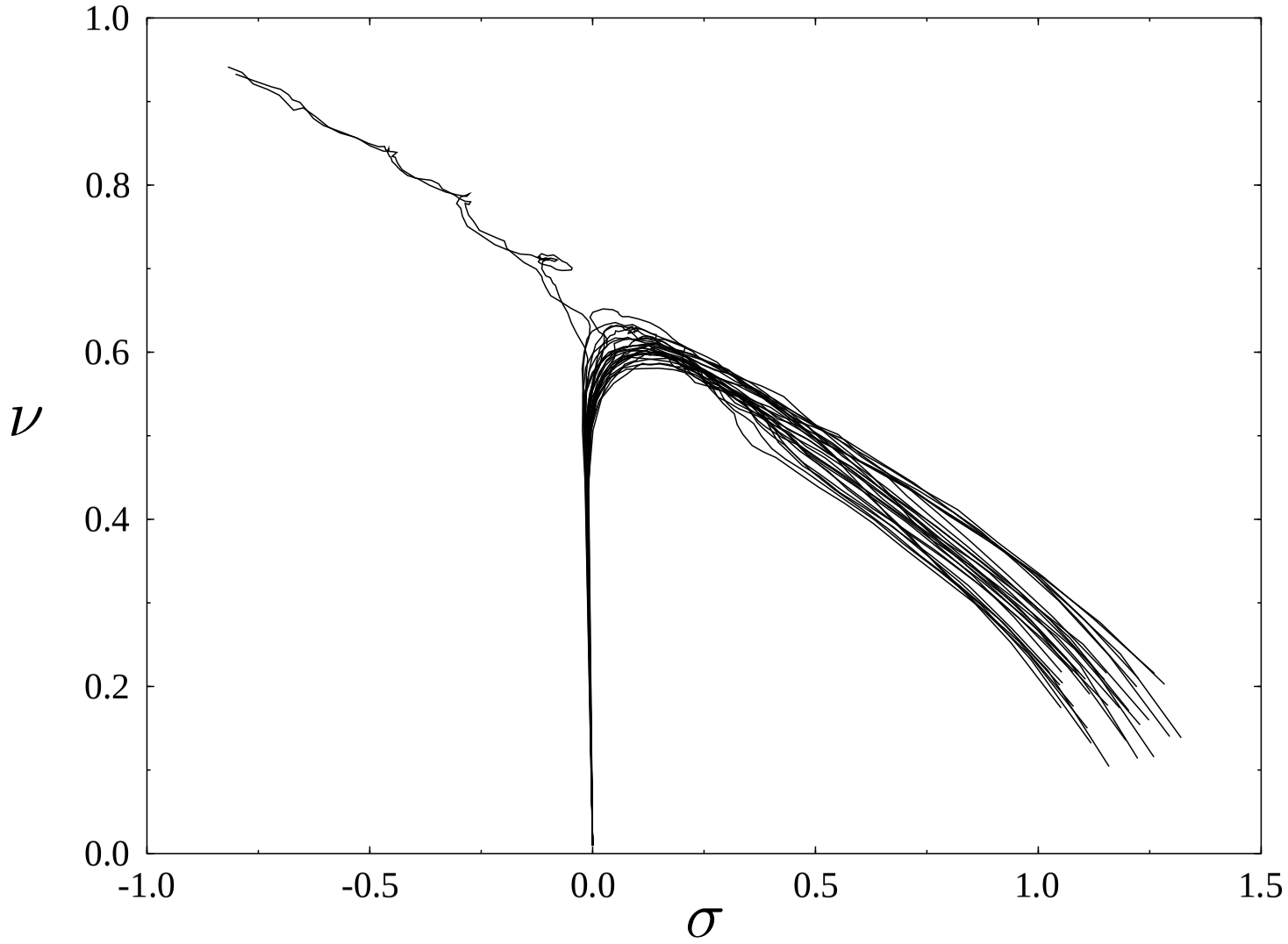
FIG. 3. Convergence of the estimator $P_{\rm fus}^{(N)}$ with number of trajectories simulated (N), for both direct simulation (solid line), and using importance sampling (dashed line).

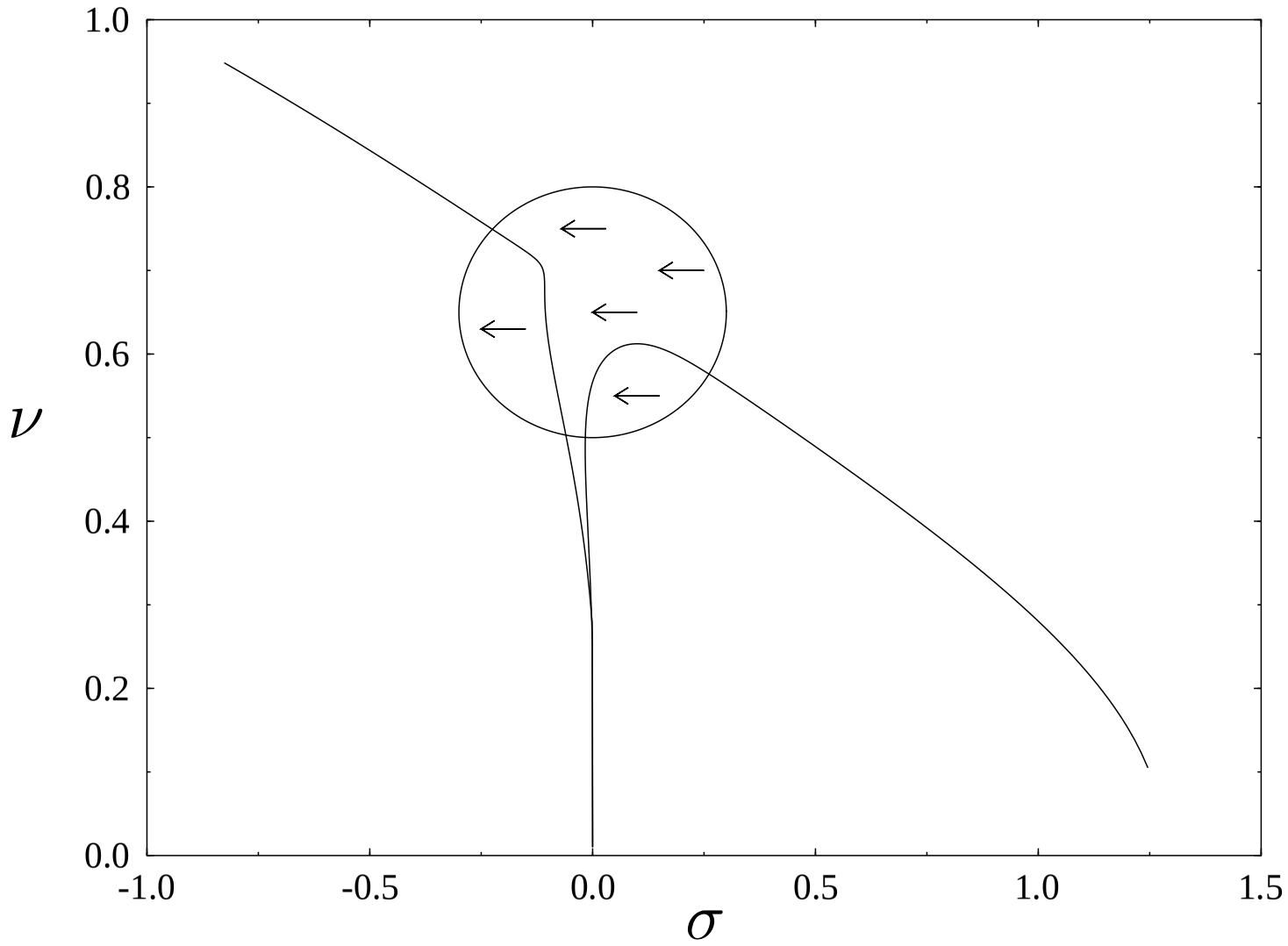
FIG. 4. Excitation function computed using direct simulation, with 1000 trajectories for each point. (The solid line is an analytical estimate extracted from a much larger number of simulations.)

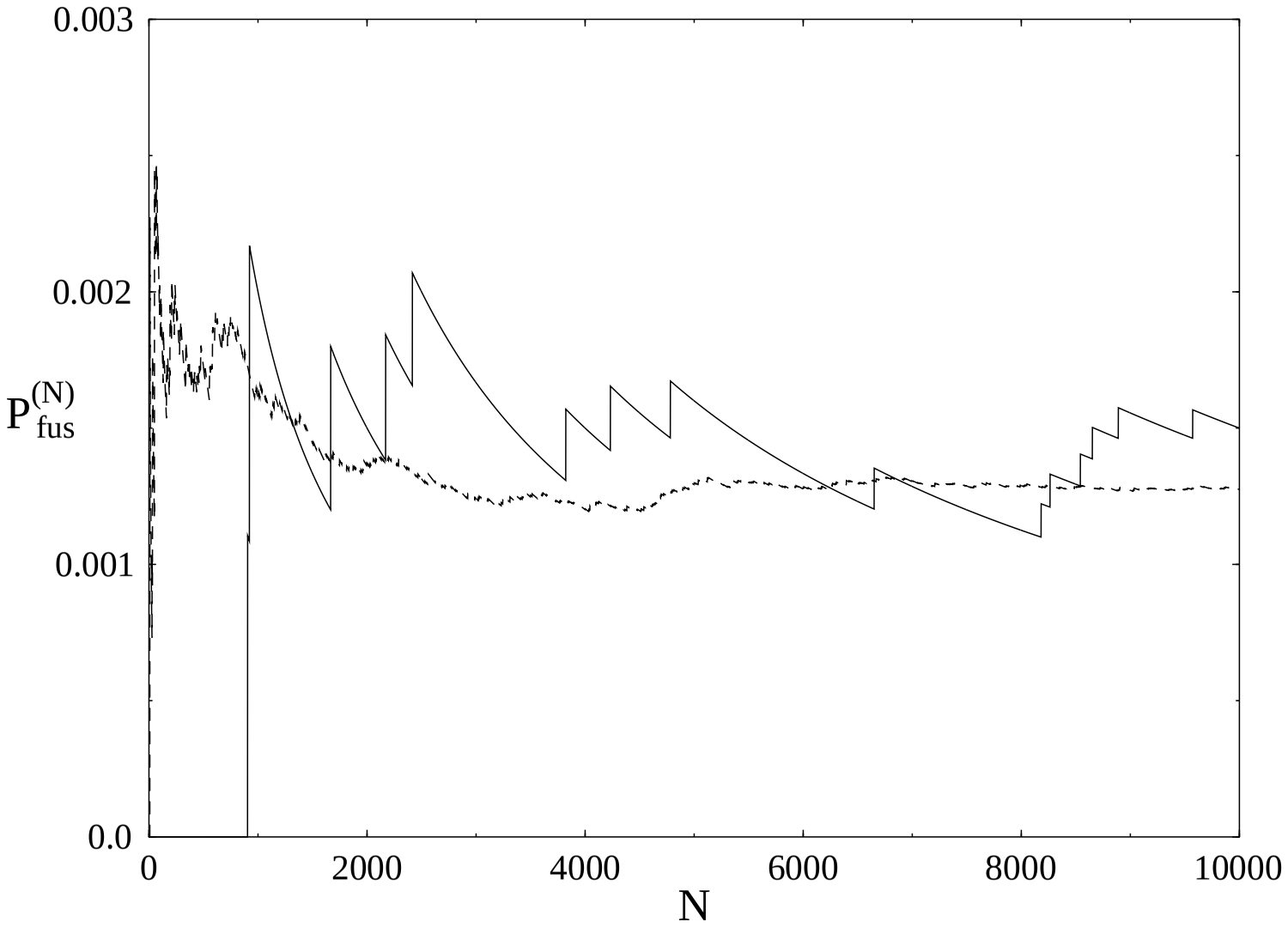
FIG. 5. Same as Fig.4, but computed using importance sampling instead of direct simulation.

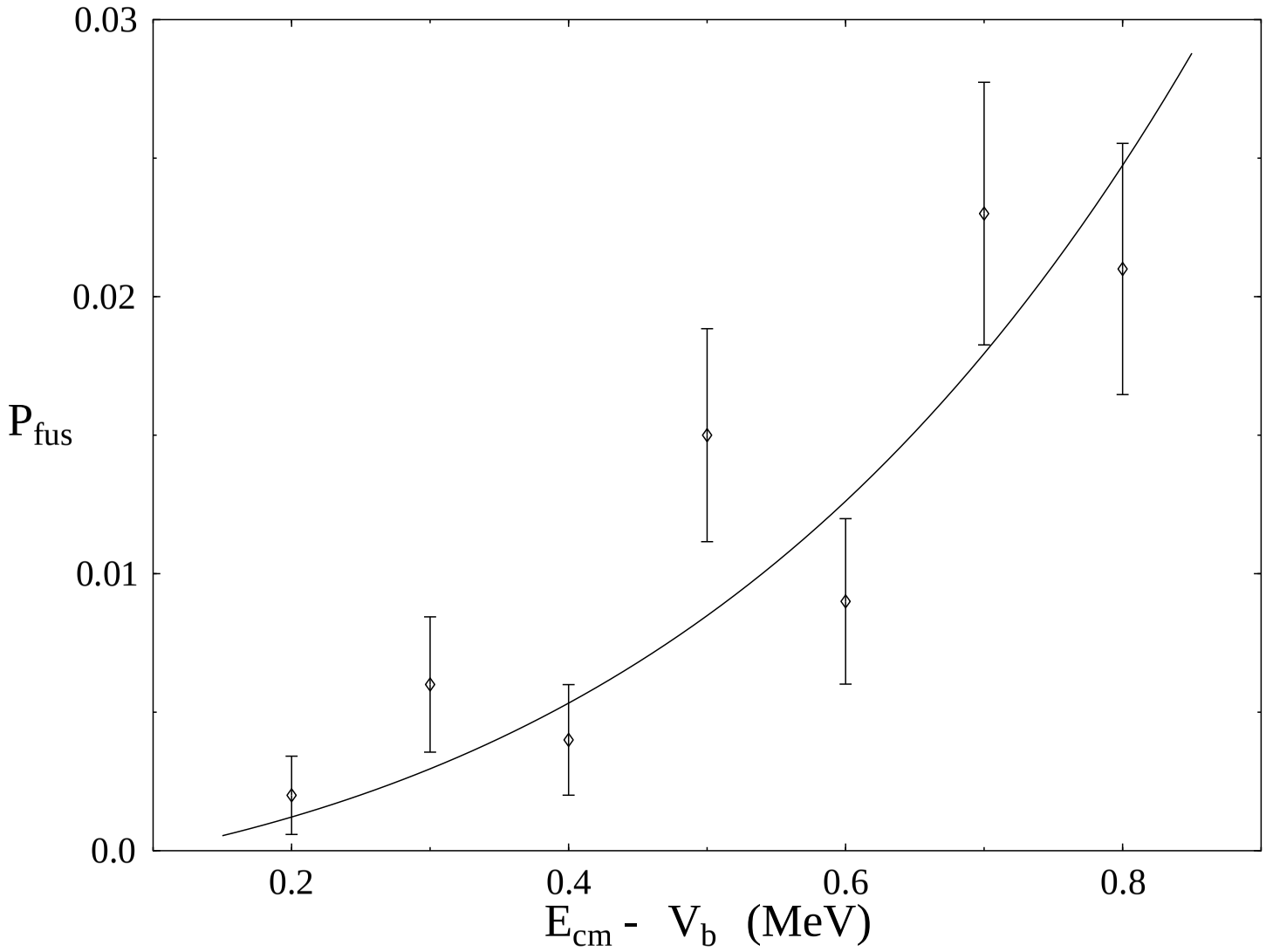
FIG. 6. Efficiency gain as a function of wind strength, Eq.51, for $x_0 = 3$ ($P = 1.35 \times 10^{-3}$). At the optimal wind value, the gain is around 220.

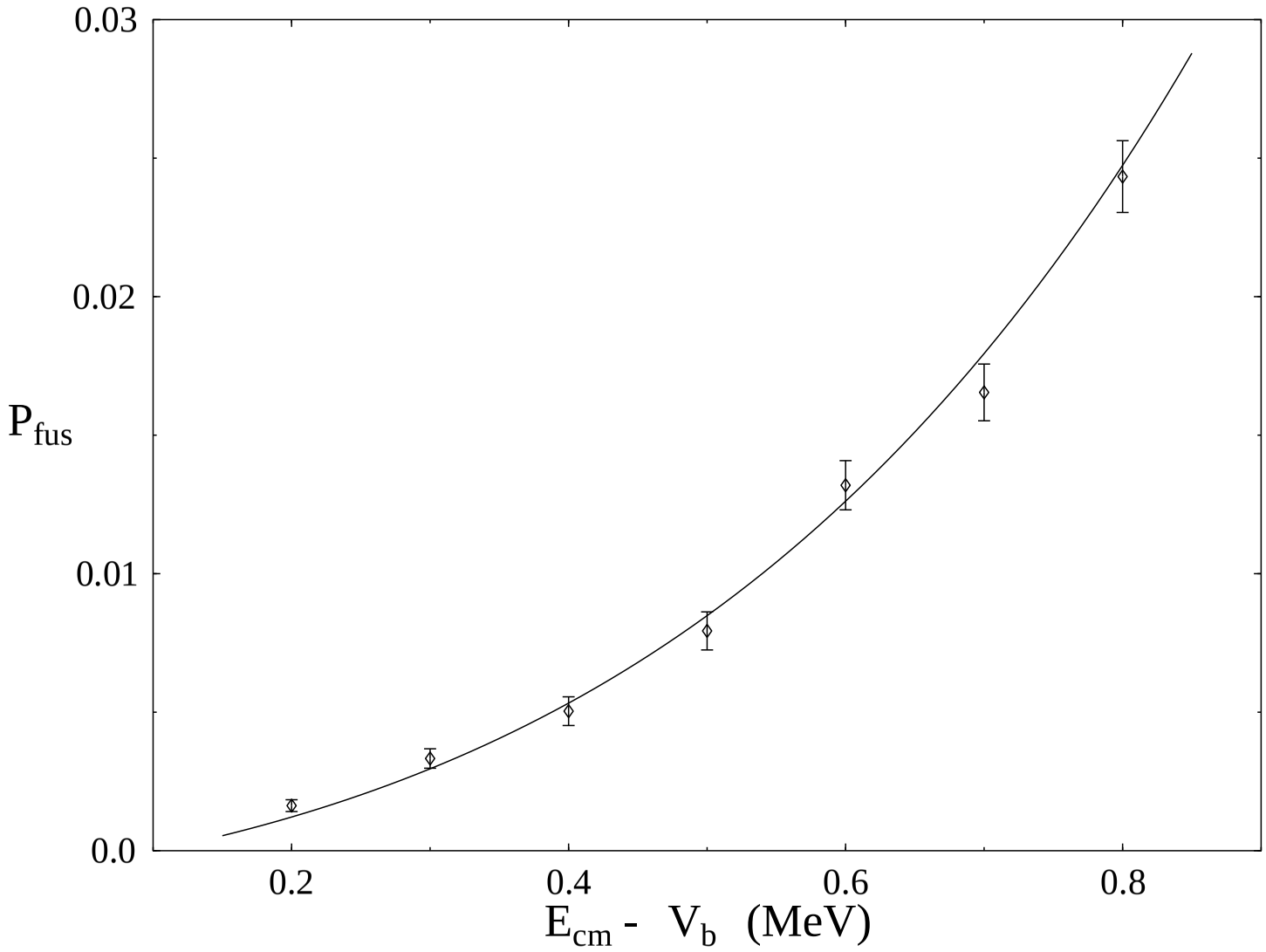
FIG. 7. Same as Fig.6, but for $x_0 = 0$. Here, $P = 0.5$, so there would be no problem in estimating this probability from direct simulations. Nevertheless, there is an efficiency gain of nearly two, when using importance sampling.

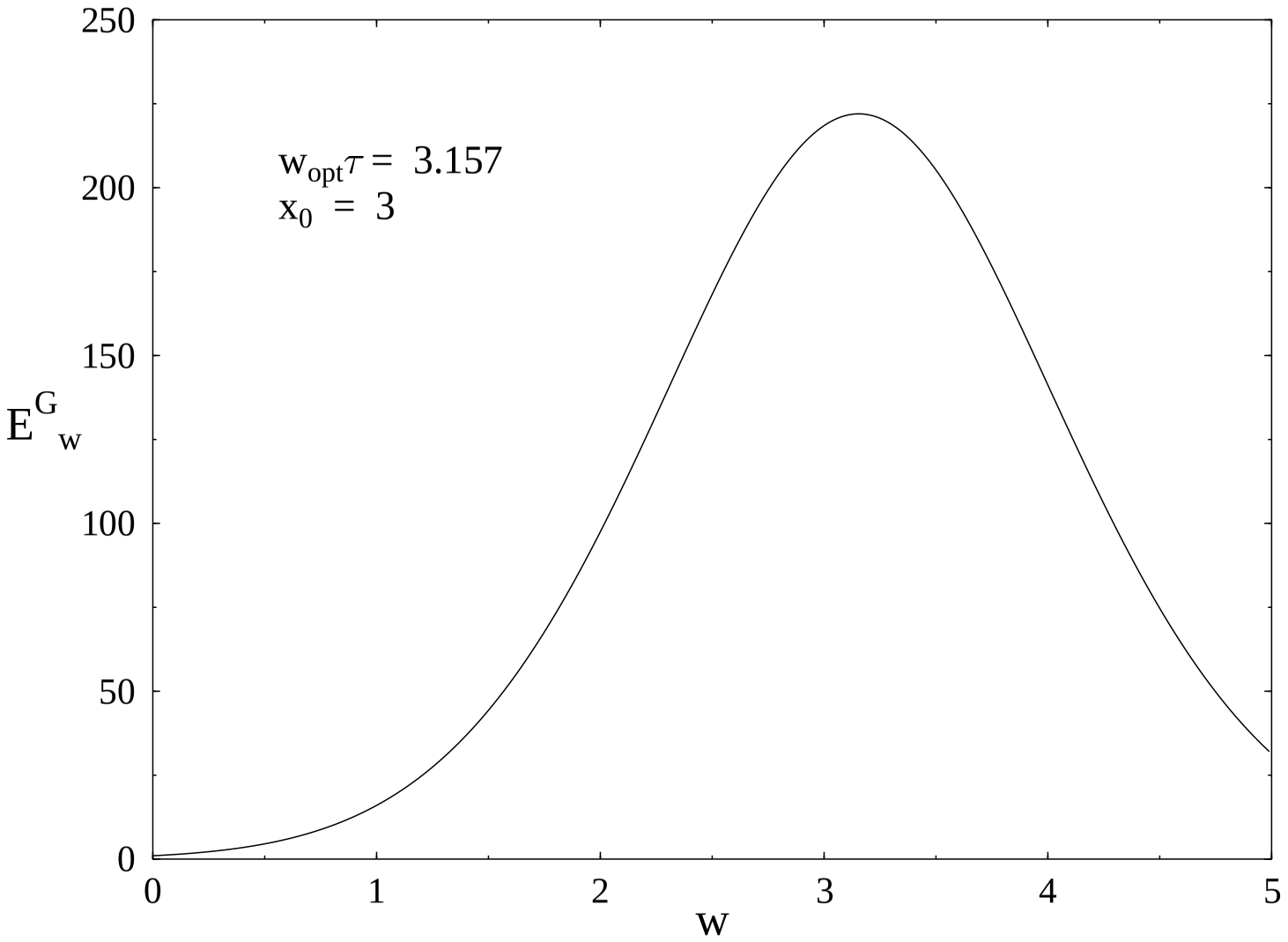

1.0
0.8
0.6
0.4
0.2
0.0
ν
-1.0
-0.5
0.0
0.5
1.0
1.5
σ

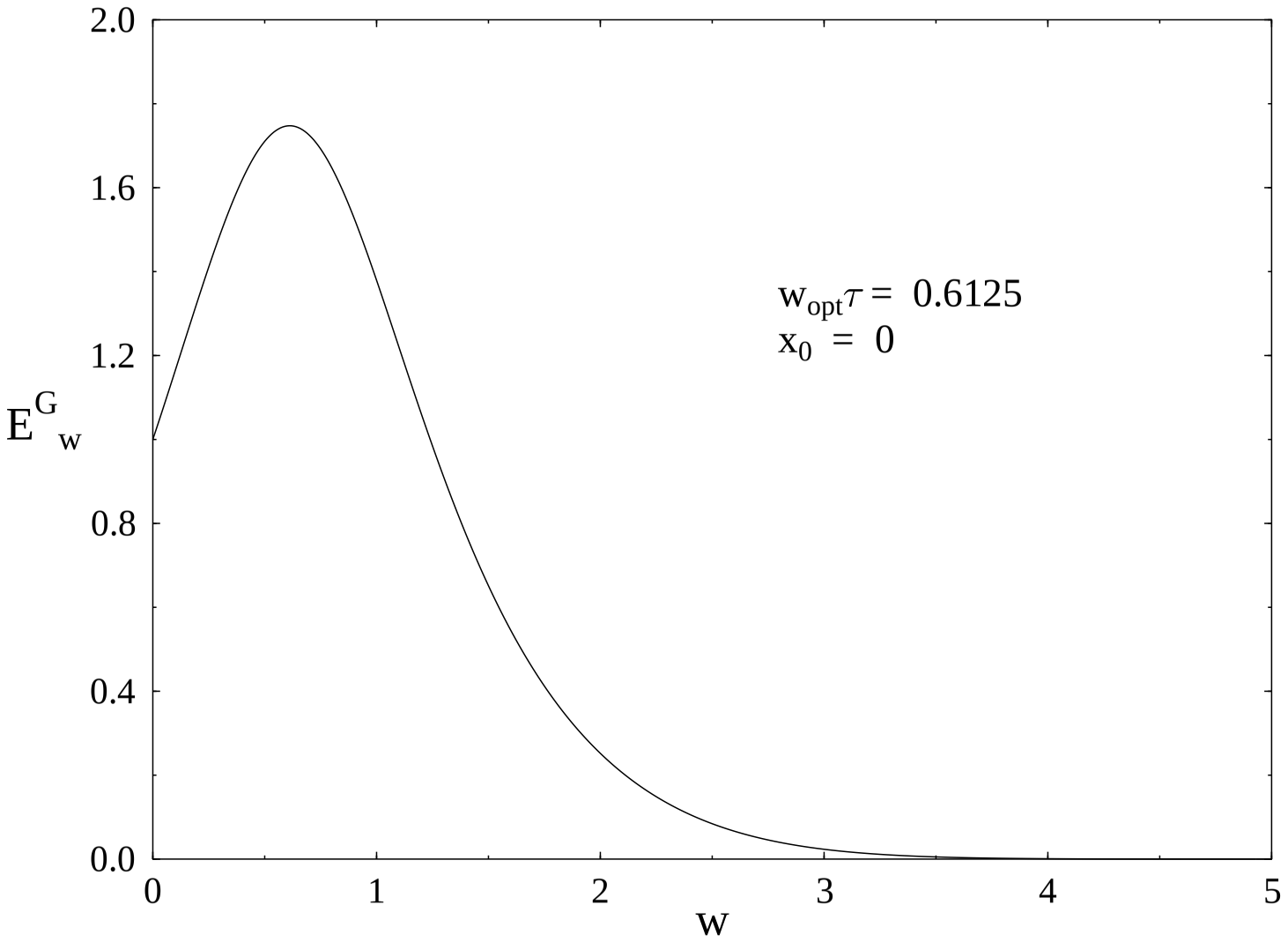

1.0
0.8
0.6
0.4
0.2
0.0
ν
-1.0
-0.5
0.0
0.5
1.0
1.5
σ


0.003
0.002
0.001
0.0
$P_{fus}^{(N)}$
0
2000
4000
6000
8000
10000
N


0.03
0.02
0.01
0.0
P_{fus}
0.2
0.4
0.6
0.8
E_{cm} - V_b (MeV)


0.03
0.02
0.01
0.0
P_{fus}
0.2
0.4
0.6
0.8
E_{cm} - V_b (MeV)


250
200
150
100
50
0
0
1
2
3
4
5
w
E^G_w
$w_{opt}\tau = 3.157$
$x_0 = 3$


2.0
1.6
1.2
0.8
0.4
0.0
E^G_w
0
1
2
3
4
5
w
$w_{opt}\tau = 0.6125$
$x_0 = 0$